

Metadata

algumas codificações

LICÍNIA SANTOS

RESUMO Fala-se muito de Metadata, aparentemente como sendo uma nova questão, resultante da aplicação das novas tecnologias às bibliotecas, nomeadamente com o aparecimento das chamadas bibliotecas digitais e/ou bibliotecas digitalizadas. Pretende-se com este artigo desmistificar o fenómeno Metadata como realidade recente, fazendo um paralelismo entre a sua utilização pelas bibliotecas tradicionais e a sua utilidade nas bibliotecas digitais. Refere-se ainda o Dublin Core como estrutura *standard* de Metadata, para identificar documentos em suporte digital e/ou na Internet.

ABSTRACT We hear a lot about Metadata as being a new issue in result of the application of new technologies to libraries, namely to digital libraries. The aim of the current paper is to dismount the Metadata issue taken as a new issue, making a parallelism between its application by traditional libraries and its usefulness in digital libraries. The Dublin Core is also referred as a Metadata standard structure to identify documents either available on digital media or in Internet.

Introdução

Fala-se muito de Metadata, aparentemente como sendo uma nova questão, resultante da aplicação das novas tecnologias às bibliotecas, nomeadamente com o aparecimento das chamadas bibliotecas digitais e/ou bibliotecas digitalizadas. Pretende-se com este artigo desmistificar o fenómeno Metadata como realidade recente, fazendo um paralelismo entre a sua utilização pelas bibliotecas tradicionais e a sua utilidade nas bibliotecas digitais. Refere-se ainda o Dublin Core como estrutura *standard* de Metadata, para identificar documentos em suporte digital e/ou na Internet.

1. A utilização de Metadata nas bibliotecas tradicionais

Em sentido mais lato, Metadata significa dados sobre os dados, isto é, um conjunto de atributos ou elementos necessários para identificar cada documento de uma determinada colecção ou fundo de uma biblioteca. Também num sentido mais abrangente, a função primeira de qualquer biblioteca é organizar a sua colecção e estruturar os dados de forma a permitir o acesso aos documentos por parte dos utilizadores. Nesta medida, podemos dizer que os catálogos tradicionais já encerram um conjunto de Metadata ou Meta-informação importante sobre os documentos. Neste tipo de Metadata incluem-se os dados resultantes da catalogação do documento segundo as ISBD's e RPC, no caso português. Representa-se pela catalogação quando se descreve um título, um complemento de título, um autor, um local de publicação, um editor, uma data de publicação, um assunto ou outros tantos elementos descritivos, que identifiquem um dado documento.

Com a transferência dos catálogos manuais para sistemas automatizados esta questão da Metadata, está cada vez mais implícita, quer pela necessidade de normalização necessária para poder disponibilizar informaticamente a informação, quer pelo facto de se poder trocar essa mesma informação entre as várias entidades que a produzem. Em Portugal, as questões normativas, no que respeita à descrição catalográfica dos documentos em sistemas automatizados passam pela utilização do UNIMARC. Assim, podemos dizer que, também o UNIMARC é um sistema de codificação de Metadata, tal como todos os restantes formatos MARC ou outros formatos para sistemas automatizados. Relativamente a esta questão e, já que estamos a referir a automatização, apenas temos de ter em atenção que a normalização é importante, sob pena de não comunicarmos uns com os outros.

A ligação de Metadata a um determinado documento pode fazer-se de duas formas distintas:

- Elementos ou atributos existentes num registo separado do item bibliográfico;
- Elementos ou atributos integrados no próprio documento.

Embora possa parecer que a primeira situação diz respeito à forma de organização da informação nas bibliotecas tradicionais e que a segunda situação diz respeito à organização do conhecimento nas bibliotecas digitais, tal não é verdade. De facto, nas bibliotecas tradicionais já se convive com estes dois tipos de organização da informação, senão vejamos: quer nos catálogos manuais, quer nos catálo-

gos automatizados, temos uma situação de organização da informação em que os Metadados estão dissociados do item. Mas, também temos situações em que essa Meta-informação se encontra integrada no próprio documento, como é exemplo da catalogação CIP (Cataloging In Publication), da ficha técnica que por vezes se apresenta nas publicações, dos elementos que se encontram na página de rosto, muitas vezes com uma disposição própria que melhor permite a sua identificação.

2. A Metadata nas bibliotecas digitais

No que respeita às bibliotecas digitais é comum conceber que este tipo de bibliotecas tem como característica principal, a gestão das suas colecções apenas em formato digital ou que, o acesso à informação se faz essencialmente recorrendo à tecnologia digital. O acesso neste tipo de bibliotecas, embora assente na estrutura de Metadata traz, uma mais valia significativa no que respeita ao acesso imediato ao documento e a possíveis conexões com outros documentos similares ou correspondentes ao assunto seleccionado. Não há, portanto, dúvidas que o acesso imediato ao texto integral e a vários documentos em simultâneo, marca a grande diferença relativamente às bibliotecas tradicionais.

Este conceito surge, intimamente, associado à Internet e, intrinsecamente ligado à «explosão» da produção de documentos electrónicos e ao aparecimento de bibliotecas digitais. Neste contexto, a Metadata é muito importante para identificar, descrever os documentos e criar a sua identidade de forma a evitar duplicações e tornar único cada documento. Estes atributos ou elementos, normalmente, aparecem integrados no próprio documento, numa estrutura própria que serve os princípios da disponibilização da informação na Internet, contexto mais comum em que se inserem as bibliotecas digitais. Desde o início da produção deste tipo de documentos que, os seus próprios autores/produtores, identificam os seus documentos com informações respeitantes à identificação do documento (título, nome do autor, data de criação, etc.). Foi este tipo de identificação, isto é, Metadata que, permitiu até então a pesquisa na Web por parte dos motores de busca. No entanto, a recuperação da informação tem sido dificultada pela forma como se apresenta a Metadata, dada a falta de uniformização na apresentação dos dados. Começamos por um conjunto de dados integrados na linguagem HTML que orienta o modo como a informação é disponibilizada na Internet, por vezes de difícil identi-

ficação por parte do bibliotecário que quer representar a informação, ou por parte do utilizador que pretende identificar o documento que tem em mãos.

Numa pesquisa de um qualquer assunto na Internet, a pertinência dos resultados faz-se, muitas vezes, tendo em consideração a entidade que produz a informação, a data de produção, data de actualização, a relação com outros documentos congéneres, etc. pelo que, a não identificação rápida deste tipo de Metadata traz alguns problemas na recuperação dos documentos. Uma outra desvantagem, da integração da Metadata na linguagem HTML (HyperText Markup Language) tem a ver com a necessidade de se entrar na programação sempre que se verifiquem alterações ao conteúdo do documento, a fim de as poder representar. Esta situação está a ser ultrapassada com a utilização do XML (eXtensible Markup Language) que, há quem designe por linguagem do futuro, por possibilitar a codificação destes elementos, visíveis no conteúdo do documento. Apesar de tudo isto, sente-se ainda uma falta de organização no conjunto dos dados a codificar ou que identifiquem o documento, daí o surgimento do Dublin Core.

3. Estrutura de Metadata de acordo com as regras do Dublin Core

O Dublin Core não é mais do que um *standard* desenvolvido no seio da OCLC (Dublin Core Metadata Initiative), com tendência a ser adoptado pelo resto do mundo, a fim de designar uma semântica comum na descrição e identificação da informação em formato digital, ganhando importância à medida que se vai desenvolvendo a Internet. Reconhece-se que a consistência na criação de Metadata é «a chave do sucesso» para uma boa recuperação da informação.

O aparecimento desta nova forma de codificar dados que, já se codificavam em formatos automatizados, surge devido à proliferação dos formatos MARC, e à adopção de regras diferentes de catalogação nos diferentes países, mesmo usando um formato igual e, também, pela dificuldade de integrar os diferentes formatos MARC na linguagem SGML (Standard Generalized Markup Language).

É uma estrutura que assenta em três objectivos bem delimitados:

- Simplicidade na criação e na manutenção dos dados (pretende-se que o produtor possa preencher os vários elementos que compõem esta estrutura, de forma simples, sem necessidade de

quaisquer conhecimentos técnico-documentais, daí o não se ter escolhido nenhum *standard* já existente);

- Semântica comum (pretende-se seguir uma mesma terminologia para representar conceitos idênticos, de forma a orientar o utilizador leigo na obtenção de resultados mais específicos);
- Âmbito internacional (inicialmente foi desenvolvido em inglês mas, já foram criadas versões em muitas outras línguas).

O Dublin Core Metadata é uma estrutura *standard* que organiza os elementos descritivos que identificam um documento de acordo com uma estrutura de 15 elementos. Esta estrutura é, ainda alvo de discussão concertada a nível internacional, e coordenada por um grupo interdisciplinar de profissionais que tem por objectivo definir o *standard* tendo em atenção uma semântica única e, não esquecendo que se destina a identificar documentos World Wide Web. Cada elemento é opcional e repetível.

3.1 Estrutura do Dublin Core

Título – nome atribuído ao documento, nome pelo qual o documento é formalmente conhecido.

Produtor – entidade primária responsável pelo conteúdo do documento (pode ser uma pessoa física, uma organização ou um serviço).

Assunto – termo de indexação que designa o conteúdo do documento (normalmente pode ser expresso por palavras-chave, frases ou através de uma classificação categorial).

Descrição – *Abstract*, índice, referência a representações gráficas, etc.

Editor – entidade responsável por disponibilizar o documento.

Colaborador – entidade responsável por prestar contribuição ao conteúdo do documento.

Data – data que especifica a vida do documento, pode ser a data de criação, de validação, data de disponibilização e/ou modificação. Deve ter um formato *standard* de oito dígitos (YYYYMMDD), conforme o recomendado pela ISO 8601.

Tipo – especifica a natureza dos dados. Normalmente inclui termos que descrevem funções ou agregações relativas ao conteúdo do documento. Também deve ser descrita aqui a forma como vão ser apresentados os dados, nomeadamente, no que respeita ao formato (de acordo com as normas emanadas por DCMI (Dublin Core Metadata Initiative) *Type Vocabulary*).

Formato – identifica o formato físico ou digital do documento. Normalmente, deve indicar-se a extensão (dimensão) do ficheiro e o tipo dos dados. Pode ainda usar-se para especificar o tipo de *software*, *hardware* ou qualquer outro equipamento necessário para visualizar o documento.

Identificador – indica-se uma referência que identifique de forma inequívoca o documento. Identificações atribuídas por diferentes sistemas, conforme o tipo de documento, como por exemplo: URI (Uniforme Resource Identifier), URL (Uniform Resource Locator), DOI (Digital Object Identifier), ISBN (International Standard Book Number), etc.

Fonte – referência à fonte do documento.

Língua – língua do texto do documento (deve ser usada uma codificação, de acordo com o definido na ISO 639)

Relação – referência a documentos relacionados. Recomenda-se que a referência seja feita de acordo com os identificadores dos documentos referenciados, como por exemplo URL. Podem ser também indicações do tipo: versão de, substituído por, parte de, referência de, etc.

Cobertura – localização espacial (cidade, país ou quaisquer coordenadas geográficas), período temporal (data) ou jurisdição (entidade administrativa). Recomenda-se a utilização de vocabulário controlado (Por exemplo *Thesaurus of Geographic Names* (TGN) ou as normas referenciadas na ISO 3166)

Direitos – informação acerca de direitos de autor (*copyright*)

3.2 *A correlação do Dublin Core com outras codificações de Metadata*

Analisando os 15 elementos atrás definidos, podemos verificar que não são muito diferentes, na sua essência, daqueles que já usamos para identificar os documentos que descrevemos e dos quais automatizamos as suas referências bibliográficas. O objectivo do Dublin Core é criar uma estrutura organizada da informação de modo a permitir aos motores de busca uma maior sistematização e ao utilizador, uma pesquisa mais fiável. É importante estruturar a informação de modo a que se, se pretender pesquisar artigos sobre braços (*arms*), não nos venha como resultado a obra técnico-científica de William Arms, um especialista na área das bibliotecas digitais. Utilizando-se esta estrutura de Metadata proposta, os motores de busca podem não ter em linha de conta as palavras do atributo Autor, em pesquisas

como a referida anteriormente, fazendo apenas referência às palavras no autor se o utilizador assim o indicar.

No mundo das bibliotecas digitais, a orientação do utilizador é um problema fulcral, mais do que nas bibliotecas tradicionais, uma vez que o contacto com o bibliotecário se perde, deixa de existir a interacção que possibilitava, muitas vezes, novas orientações no trabalho desenvolvido pelo bibliotecário. Assim, como em qualquer trabalho que se desenvolva, é necessário avaliar constantemente os resultados alcançados de modo a proceder aos ajustes necessários.

Relativamente ao Dublin Core, há quem defenda uma maior semelhança no preenchimento destes elementos de Metadata e aqueles que se utilizam numa qualquer codificação do tipo MARC. Tendo em atenção que o Dublin Core foi criado para ser preenchido pelo produtor do documento, há também opiniões que defendem que um preenchimento mais minucioso não é funcional.

A título conclusivo, é importante não esquecermos como a normalização é definitiva, tanto mais quando se trata de informatização de catálogos, de bibliotecas digitais e da disponibilização de informação na Internet. Sendo assim, e, não sendo esta matéria uma completa novidade como ficou demonstrado, aqui fica um desafio – porque não aproveitar os mesmos dados, para as diferentes codificações ou estruturas de Metadata, utilizando-os conforme as necessidades e evitando ao máximo as duplicações de trabalho.

BIBLIOGRAFIA

BECKETT, Dave; MILLER, Eric; BRICKLEY, Dan

Using Dublin Core in XML <http://purl.org/dc/documents/wd/demes-xml-20000714.htm>

Dublin Core Metadata Initiative : Making it easier to find information

<http://dublincore.org>

Dublin Core Qualifiers

<http://purl.org/dc/documents/rec/demes-qualifiers-20000711.htm>

FOX, Edward A.; SORNIL, Ohm

Digital Libraries. Draft prepared to appear as *Chapter 11 in Modern Information Retrieval, AWL England, 1999*

HILLMANN, Diane

Using Dublin Core

<http://purl.org/dc/documents/wd/usageguide-20000716.htm>